Semantic Technologies Applied to FOIA Review

William Underwood

Partnerships in Innovation: Serving a Networked Nation
November 15-16, 2004

Archival Review

- The Freedom of Information Act
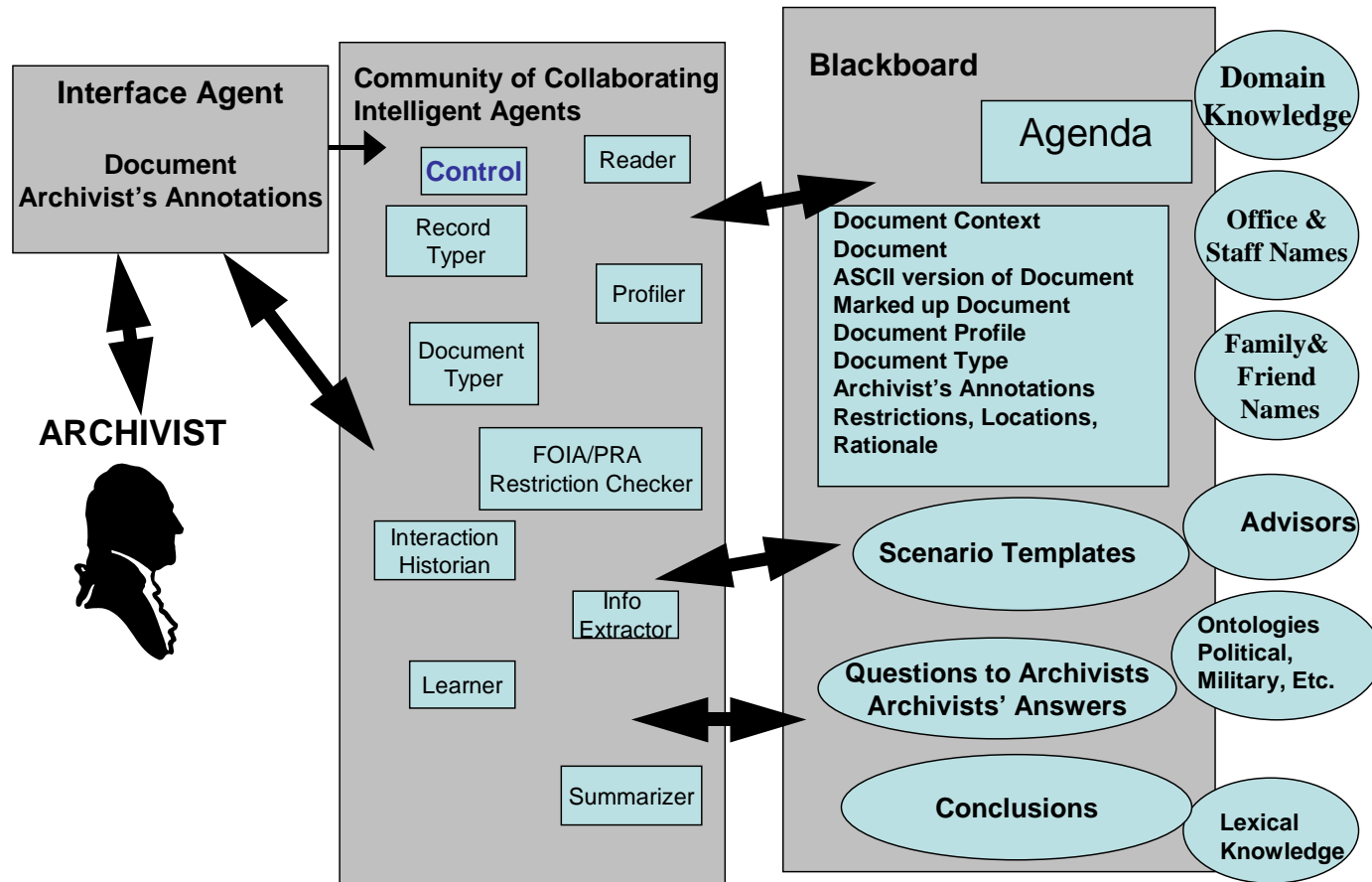- Presidential Records Act

FOIA and PRA Access Restrictions

a(1), b(1) national security and foreign policy

a(2) appointments to Federal offices

a(3) b(3) exempted by statute

a(4) b(4) confidential commercial information

a(5) confidential advice

a(6) b(6) personal privacy

b(2) personnel rules and practices of an agency

b(5) deliberative process privilege

b(7) law enforcement investigations

b(8) financial institution reports

b(9) geological information about wells

The FOIA and PRA Review Problem

- Review is an intellectually demanding task

- Requires page-by-page review

- An increasing volume of Presidential electronic records

- Limited human resources that can be applied.

- The review process is an archival processing bottleneck

Access Restriction Checker

Access Restriction Architecture

Relevant Semantic Technologies

- Information Extraction

- Content Extraction

- Knowledge Representation

- Ontologies

- Software Agents

Information Extraction

- *Information extraction* (IE) is a procedure that selects, extracts and combines data from text in order to produce structured information.

- *Named entity task* is to identify all named persons, organizations, locations, dates, times, numeric monetary amounts and percentages in text.

Other Information Extraction Tasks

- TE (Template Element) Can templates about persons and organizations be filled from an automatic analysis of text?

- CO (Co-reference) Can co-referring noun phases in text be identified, tagged and linked?

- ST (Scenario Templates) Can templates about events and their participants (persons, organizations, etc.) be filled from an automatic analysis of text?

Letter From George Bush to Ronald Reagan

```
<?xml version="1.0" encoding="UTF-8"?>
<DOCUMENT>
<P>
January 11, 1992
```

Dear Ron,

   I am writing this to you on my new computer at Camp David.
When I got back from throwing up on our friend P.M. Miyazawa,
I found that large lovely picture of the 5 of us. It will go into that beautiful
frame you gave us. Thanks so very much for such a special present.

   The trip to Asia, in spite of the carping and whining from the Democrat
protectionists, aided and abetted by the Washington Post, was a success.
We made progress getting into the glass, computer and yes, auto, markets.
Much left to be done but at least we tried and did it without reverting to
Protectionism.

   Again my sincerest thanks for that present.
I hope you are well and enjoying life to the fullest.

                                        Warmest Regards,

                                        George

```
</P>
</DOCUMENT>
```

Named Entity Recognition

```
<?xml version="1.0" encoding="UTF-8"?><DOCUMENT>
<P>
<sent>January 11, 1992


<ENAMEX id="6" type="PERSON">Dear</ENAMEX> <ENAMEX id="7" type="PERSON">Ron</ENAMEX>,

    I am writing this to you on my new computer at <ENAMEX id="8" type="LOCATION">Camp</ENAMEX>
<ENAMEX id="9" type="PERSON">David.</ENAMEX></sent>
<sent>When I got back from throwing up on our friend <ENAMEX id="10" type="PERSON">P.M. Miyazawa</ENAMEX>,
I found that large lovely picture of the 5 of us.</sent> <sent>It will go into that beautiful
frame you gave us.</sent>  <sent>Thanks so very much for such a special present.</sent>

    <sent>The trip to <ENAMEX id="11" type="LOCATION">Asia</ENAMEX>, in spite of the carping and whining from the Democrat
protectionists, aided and abetted by the <ENAMEX id="12" type="ORGANIZATION">Washington Post</ENAMEX>, was a success.</sent>
<sent>We made progress getting into the glass, computer and yes, auto, markets.</sent>
<sent>Much left to be done but at least we tried and did it without reverting to
<ENAMEX id="13" type="LOCATION">Protectionism.</ENAMEX></sent>

    <sent>Again my sincerest thanks for that present.</sent>
    <sent>I hope you are well and enjoying life to the fullest.</sent>

                          <sent><ENAMEX id="14" type="PERSON">Warmest Regards</ENAMEX>,

                          <ENAMEX id="15" type="PERSON">George</ENAMEX></sent>
</P>
</DOCUMENT>
```

Named Entity Recognition

Evaluating the Accuracy of Named Entity Recognition Technology

Annotation Type	Correct	Partially Correct	Missing	Spurious	Precision	Recall	F-Measure
Person	242	94	75	107	0.65237020316	0.70316301703	0.67681498829
Location	249	14	111	9	0.94117647058	0.68449197860	0.79256965944
Organization	146	105	59	65	0.62816455696	0.64032258064	0.63418530351
Date	162	3	30	6	0.95614035087	0.83846153846	0.89344262295
Money	5	0	0	0	1.0	1.0	1.0
Percent	9	0	4	0	1.0	0.69230769230	0.81818181818

Content Extraction Applied to Recognizing Request for Confidential Advice

December 5, 1990

MEMORANDUM FOR BOYDEN GRAY
THROUGH: BRENT SCOWCROFT
FROM: THE PRESIDENT

Boyden —

Please prepare for me a short analysis of the War Powers Resolution. . . . With out recognizing the constitutional validity of the War Powers Resolution, is there a way for the President to fulfill all his responsibilities to Congress by saying, a few days before any fighting was to begin, "hostilities are imminent—period!!

I am several thousand miles south, but these questions stay on my mind:

1. How do we fully involve Congress?
2. If we have to attack from a cold start how does the latest UN Resolution impact on congress?
    1. Is there something short of "declaring" war that satisfies Congress yet doesn't risk tying the President's hands?
    2. As the clock on the UN resolution keeps running toward the time when force has international authority, what possible official requests can/should a President make of Congress?

If you reply to this memorandum before I return, please hand carry your reply to Brent for "Eyes Only" transmission to me.

Please share a copy of this memorandum with John Sununu and Brent only.

Warm regards.

Content Extraction and Access Restriction Rules

Template(X)

Action: Request

Agent: Person

Job_Title: President

Object: Confidential Advice

Patient: C Boyden Gray

Job_Title: Counsel to the President

Presidential_Advisor: C Boyden Gray

If Document(X), and

Action(X) = Request, and

Agent(X) = Y, and

(Job_Title(Y) = President, or Presidential_Advisor(Y)) and

Patient(X) = Z and

Presidential_Advisor(Z) and

Object(X) = Confidential Advice

Then Access_Restriction(X) = a(5).

Co-reference in a Document

Start	End	ANNIE response	Annotation Features
448	456	Mr. Bush	{rule=PersonFinal, gender=male, rule1=PersonTitle}
151	171	Chairman Lee Atwater	{rule=PersonTitle1, gender=male, matches=[471, 484, 485, 486, 487]}
461	472	Mr. Atwater	{rule1=PersonTitle, rule=PersonFinal, matches=[471, 484, 485, 486, 487], gender=male}
580	583	Lee	{kind=PN, rule=Unknown, matches=[471, 484, 485, 486, 487], NMRule=Unknown}
77	91	President Bush	{rule1=PersonTitle, rule=PersonFinal}
39	46	Atwater	{kind=PN, rule=Unknown, matches=[471, 484, 485, 486, 487], NMRule=Unknown}
35	38	Lee	{gender=male, matches=[471, 484, 485, 486, 487], NMRule=Unknown, kind=PN, rule=Unknown}

Some Document Types in Bush Presidential Electronic Records

- Agenda
- Biographical Information
- Briefing Memo
- Decision Memo
- Executive Order
- Information Memo
- White House Letter
- List of Candidates for Appointment to Federal Office
- Mailing List
- Minutes of Meeting
- Nomination for Appointment to Federal Office
- Press Release
- Resume
- Schedule
- Telephone Call Recommendation

Document Type Recognition

- Convert document format to ASCII or HTML

- Use Information Extraction Technology to Markup Different Document Types.

- Machine Learning of Document Type

- Evaluate Performance

- Use for Recognizing Document Types of other Records

Other Research in Applying Semantic Technologies to Electronic Archives

- **Archival Description**

- **Response to FOIA requests**

- **High Degree of Recall and Precise Access to Records in a Very Large Collections.**

Additional Information

- http://perpos.gtri.gatech.edu

- Archival Processing Tools: User Manual

- An Analysis of the Knowledge Required to Perform FOIA and PRA Review, PERPOS Technical Report ITTL/CSITD 04-1,Mar 2004.

- PERPOS: Results of Laboratory Experiments and Use by Archivists, Nov 2003

- Recognizing Named Entities in Presidential Electronic Records, PERPOS Technical Report ITTL/CISTD 04-4, June, 2004